Robust Principal Component Analysis

Report by Akshat Doshi
SEAS-AU
Roll No: 1401119

***Abstract*—Principal Component Analysis (PCA) is the most widely useful statistical technique. One drawback of the Principal Component Analysis (PCA) method is that it uses least squares estimation technique and hence it is very sensitive to "outliers" because one single corrupted point will completely result in gross error.In computer vision applications, outliers typically occur within a sample due to some noise, errors, or occlusion. This paper suggests that a substantial amount of low rank matrices, which cannot be recovered by standard robust PCA, become recoverable by solving a convenient convex program called Principal Component Pursuit(PCP). The result theoretically justifies the effectiveness of features in robust PCA.**

Keywords: Principal component, principal component pursuit, low rank matrix, sparse matrix, video surveillance

I. Introduction

Robust principal component analysis receives much attention in recent studies for its ability to recover the low rank model from sparse noise. Such sparse structure of noise is common in many real time applications such as face recognition, video surveillance, latent semantic indexing, ranking and collaboration filtering, system identification, graphical modeling. A key application where this occurs is in video analysis where we have given video frames as data matrix and we have to separate a background from moving foreground objects. The main agenda behind doing this is suppose we have data matrix M which is a superposition of two matrices low rank(L) and sparse(S). So, the goal here is to separate the L and the S component from a corrupted matrix M.

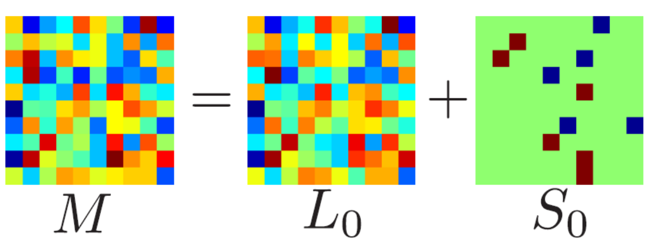

Fig. 1: Separation problem

II. The separation problem

Suppose we are given a large data matrix $M \in R^{n_1 \times n_2}$ which can decompose into a lower rank matrix L and a sparse matrix S as,

$$M = L + S \tag{1}$$

We do not have any information about low rank and sparse matrix still we have to reconstruct both low rank (L) and sparse (S) matrix with the help of data matrix (M). Robust PCA is NP-hard problem. Since, we do not know how to solve un convex NP-hard problem we relax it to convex optimization problem.

III. Convex optimization problem

Robust PCA's aim is to separate L and S from given data matrix M. This problem can be solved by convex optimization. So we would like to find the best fit for L and S matrix such that their sum generates a perfect data matrix M and among all this decomposition we would like to find one with minimum complexity. The principal component pursuit (PCP) propose as,

$$\begin{aligned} &\text{minimize} \quad && ||L||_* + \lambda ||S||_1 \\ &\text{subject to} \quad && L + S = M \end{aligned}$$

Let $||L||_* := \sum_i \sigma_i(L)$ denote the nuclear norm of low rank matrix L that is the sum of the singular value of L, and let $||S||_1 := \sum_{ij} |S_{ij}|$ denote the l_1 norm of matrix S seen as sum of absolute values of element S. If data entry corrupts our convex optimization algorithm will definitely recover the data back but under certain conditions. So now we will look at all of that assumption that we have derived to get perfect outcome.

A. Assumptions

- One assumption is that low rank matrix should not be sparse matrix and vise-versa.
  Suppose the matrix M is equal to $e_1 e_1^*$ (this matrix has a one in the top left corner and zeros everywhere else). So we can not decide whether M is low-rank or sparse matrix.
  To resolve this problem paper introduces notion of incoherence. It states that if low rank component is very sparse than the problem is hard to do. So we resolve this geometrically. Therefore, we have low rank component L, by computing SVD we get column space and row space. So we have matrix L whose rank is r and to find incoherence firstly we perform SVD on matrix $L \in R^{n_1 \times n_2}$,

$$L = U \Sigma V^* = \sum_{i=1}^{r} \sigma_i u_i v_i$$

  Where Σ is a matrix with singular values as its diagonal entries, U is a matrix with columns as the left singular vectors of L, V is a matrix with columns as the right singular vectors of L, and V^* denotes the transpose of V. Now, we measure the correlation of row space and column space along with its basis vector. So condition for that is stated as,

$$\max_i ||U^* e_i||^2 \leq \frac{\mu r}{n_1}, \qquad \max_i ||V^* e_i||^2 \leq \frac{\mu r}{n_2} \tag{2}$$

$$||UV^*||_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}} \tag{3}$$

If the coherence(μ) is very large than situation like (L and S look similar) this occurs, but for small values of μ, the singular vectors are randomly spread out—in other words, not sparse. So to satisfy our condition, μ should always be around one.

Another identifiability issue arises if the sparse matrix has low-rank. This will occur say, if, all the nonzero entries of S occur in a column or in a few columns. For instance, the first column of S_0 is opposite to L_0, so all the other columns of S_0 vanish. Then it is clear that we would not be able to recover L_0 and S_0 by any method whatsoever since $M = L_0 + S_0$ would have a column space equal to, or included in that of L_0. To avoid such meaningless situations, we will assume that the sparsity pattern of the sparse component is selected uniformly at random.

- Second problem is we have perfect low rank and sparse component but its entries are corrupted so to recover the corrupted entries this solution perfectly recovers all the component. So this theorem suggest at what probability your entry will be corrupted. Suppose L_0 is $n \times n$, obeys (2)-(3). Fix any $n \times n$ matrix Σ of $signs$. Suppose that the support set Ω of S_0 is uniformly distributed among all sets of cardinality m, and that sgn($[S_0]_{ij}$) = $\sum_{ij}$ for all $(i, j) \in \Omega$. Then, there is a numerical constant c such that with probability at least 1 - cn^{-10} (over the choice of support of S_0),Principal Component Pursuit with $\lambda = \frac{1}{\sqrt{n}}$ is exact, that is, $\hat{L}$=L_0 and $\hat{S}$=S_0, provided that

$$rank(L_0) \leq \rho_r n \mu^{-1} (\log n)^{-2} \; and \; m \leq \rho_s n^2 \tag{4}$$

In this equation, ρ_r and ρ_s are positive constants. In the general rectangular case, where L_0 is $n_1 \times n_2$, PCP with $\lambda = \frac{1}{\sqrt{n_1}}$ succeeds with probability at least 1 - cn_1^{-10}, provided that $rank(L_0) \leq \rho_r n_2 \mu^{-1} (\log n_1)^{-2} \; and \; m \leq \rho_s n_1 n_2$.

So from above result we can say that the rank of low rank component is in terms of $n/\log(n)$ which is quit large. So every entry has positive chance to corrupt let's say 10% and now with this probability we optimize the problem with $\lambda = \frac{1}{\sqrt{n}}$ and the failure probability is in terms of n^{-b}, $b > 0$ which is minute so we can successfully retrieve our data matrix.

This paper propose different lemmas and proofs like Elimination Theorem, Derandomization, Dual Certificates, Dual Certification via the Golfing Scheme, ADMM, ALM to prove this given theorem.

B. Algorithm

For getting good accuracy and convergence across wide range of problem, the paper uses Augmented Lagrange Multiplier (ALM) to resolve convex PCP problem. ALM has more accuracy than APG and that to also in fewer iterations. The ALM algorithm also has a property that the rank of the iterates often remains bounded by rank(L_0) throughout the optimization, allowing them to be computed especially efficiently. The algorithm is given as follows:

Algorithm 1 PCP by Alternating Directions

1: **initialize**: $S_0 = Y_0 = 0, \mu > 0$
2: **while** not converged **do**
3: compute $L_{k+1} = D_{1/\mu}(M - S_k + \mu^{-1} Y_k)$
4: compute $S_{k+1} = S_{\lambda/\mu}(M - L_{k+1} + \mu^{-1} Y_k)$
5: compute $Y_{k+1} = Y_k + \mu(M - L_{k+1} - S_{k+1})$
6: **end while**
7: **output:** L,S

IV. Conclusion

We characterize refined conditions under which PCP succeeds to solve the robust PCA problem. Such result is well supported by our numerical experiments. Our result shows that the ability of PCP to correctly recover a low-rank matrix from errors is related not only to the total number of corrupted entries but also to locations of corrupted entries, more essentially to the local incoherence of the low rank matrix. one can disentangle the low-rank and sparse components exactly by convex programming, and this provably works under quite broad conditions. Further, our analysis has revealed rather close relationships between matrix completion and matrix recovery (from sparse errors) and our results even generalize to the case when there are both incomplete and corrupted entries.

V. References

[1]Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization. John Wright , Yigang Peng, Yi Ma Visual Computing Group Microsoft Research Asia

[2]Analysis of Robust PCA via Local Incoherence. Huishuai Zhang, Yi Zhou, Yingbin Liang.

[3]Robust Principal Component Analysis? by EMMANUEL J. CANDE'S and XIAODONG LI, Stanford University YI MA, University of Illinois at Urbana-Champaign, Microsoft Research Asia JOHN WRIGHT, Microsoft Research Asia. code implementation https://github.com/dlaptev/RobustPCA

[4]http://perception.csl.illinois.edu/matrix-rank/home.html